Gaussian random projections for Euclidean membership problems

Vu Khac Ky[1], Pierre-Louis Poirion, Leo Liberti

CNRS LIX, École Polytechnique, F-91128 Palaiseau, France
Email:vu,poirion,liberti@lix.polytechnique.fr

November 19, 2015

Abstract

We discuss the application of random projections to the fundamental problem of deciding whether a given point in a Euclidean space belongs to a given set. We show that, under a number of different assumptions, the feasibility and infeasibility of this problem are preserved with high probability when the problem data is projected to a lower dimensional space. Our results are applicable to any algorithmic setting which needs to solve Euclidean membership problems in a high-dimensional space.

1 Introduction

Random projections are very useful dimension reduction techniques which are widely used in computer science [7, 13]. We assume we have an algorithm $\mathcal{A}$ acting on a data set X consisting of n vectors in $\mathbb{R}^m$, where m is large, and assume that the complexity of $\mathcal{A}$ depends on m and n in a way that makes it impossible to run $\mathcal{A}$ sufficiently fast. A random projection exploits the statistical properties of some random distribution to construct a mapping which embeds X into a lower dimensional space $\mathbb{R}^k$ (for some appropriately chosen k) while preserving distances, angles, or other quantities used by $\mathcal{A}$.

One striking example of random projections is the famous Johnson-Lindenstrauss lemma [9]:

1.1 Theorem (Johnson-Lindenstrauss Lemma)
Let X be a set of m points in $\mathbb{R}^m$ and $\varepsilon > 0$. Then there is a map $F : \mathbb{R}^m \to \mathbb{R}^k$ where k is $O(\frac{\log m}{\varepsilon^2})$, such that for any $x, y \in X$, we have

$$(1-\varepsilon)\|x-y\|_2^2 \le \|F(x)-F(y)\|_2^2 \le (1+\varepsilon)\|x-y\|_2^2. \tag{1}$$

Intuitively, this lemma claims that X can be projected in a much lower dimensional space whilst keeping Euclidean distances approximately the same. The main idea to prove Thm. 1.1 is to construct a random linear mapping T (called *JL random mapping* onwards), sampled from certain distribution families, so that for each $x \in \mathbb{R}^m$, the event that

$$(1-\varepsilon)\|x\|_2^2 \le \|T(x)\|_2^2 \le (1+\varepsilon)\|x\|_2^2 \tag{2}$$

[1]Supported by a Microsft Research Ph.D. fellowship.

occurs with high probability. By Eq. (2) and the union bound, it is possible to show the existence of a map F with the stated properties (see [2, 4]).

In this paper we employ random projections to study the following general problem:

Euclidean Set Membership Problem (ESMP). Given $p \in \mathbb{R}^m$ and $X \subseteq \mathbb{R}^m$, decide whether $p \in X$.

This is a fundamental class consisting of many problems, both in **P** (e.g. the Linear Feasibility Problem (LFP)) and **NP**-hard (e.g. the Integer Feasibility Problem (IFP), which can naturally model SAT, and also see [15]).

In this paper, we use a random linear projection operator T to embed both p and X to a lower dimensional space, and study the relationship between the original membership problem and its projected version:

Projected ESMP (PESMP). Given p, X, T as above, decide whether $T(p) \in T(X)$.

Note that, when $p \in X$,the fact that $T(p) \in T(X)$ follows by linearity of T. We are therefore only interested in the case when $p \notin X$, i.e. we want to estimate $\mathsf{Prob}(T(p) \notin T(X))$, given that $p \notin X$.

1.1 Previous results

Random projections applying to some special cases of membership problems have been studied in [11], where we exploited some polyhedral structures of the problem to derive several results for polytopes and polyhedral cones. In the case X is a polytope, we obtained the following result.

1.2 Proposition ([11])
Given $a_1, \ldots, a_n \in \mathbb{R}^m$, let $C = \mathsf{conv}\{a_1, \ldots, a_n\}$, $b \in \mathbb{R}^m$ such that $b \notin C$, $d = \min\limits_{x \in C} \|b - x\|$ and $D = \max\limits_{1 \le i \le n} \|b - a_i\|$. Let $T : \mathbb{R}^m \to \mathbb{R}^k$ be a JL random mapping. Then

$$\mathsf{Prob}\big(T(b) \notin T(C)\big) \ge 1 - 2n^2 e^{-\mathcal{C}(\varepsilon^2 - \varepsilon^3)k}$$

for some constant $\mathcal{C}$ (independent of m, n, k, d, D) and $\varepsilon < \frac{d^2}{D^2}$.

If X is a polyhedral cone, we obtained the following result.

1.3 Proposition ([11])
Given $b, a_1, \ldots, a_n \in \mathbb{R}^m$ of norms 1 such that $b \notin C = \mathsf{cone}\{a_1, \ldots, a_n\}$, let $d = \min\limits_{x \in C} \|b - x\|$ and $T : \mathbb{R}^m \to \mathbb{R}^k$ be a JL random mapping. Then:

$$\mathsf{Prob}\big(T(b) \notin T(C)\big) \ge 1 - 2n(n+1) e^{-\mathcal{C}(\varepsilon^2 - \varepsilon^3)k}$$

for some constant $\mathcal{C}$ (independent of m, n, k, d), where $\varepsilon = \frac{d^2}{\mu_A^2 + 2\sqrt{1-d^2}\mu_A + 1}$,

$$\mu_A = \max\{\|x\|_A \mid x \in \mathsf{cone}(a_1, \ldots, a_n) \wedge \|x\| \le 1\},$$

and $\|x\|_A = \min\big\{ \sum_i \theta_i \mid \theta \ge 0 \wedge x = \sum_i \theta_i a_i \big\}$ is the norm induced by $A = (a_1, \ldots, a_n)$.

We also recall the following Lemma, useful for the integer case.

1.4 Lemma ([11])
Let $T : \mathbb{R}^m \to \mathbb{R}^k$ be a JL random mapping, let $b, a_1, \ldots, a_n \in \mathbb{R}^m$ and let $X \subseteq \mathbb{R}^m$ be a finite set. Then if $b \neq \sum_{i=1}^m y_i a_i$ for all $y \in X$, we have

$$\mathsf{Prob}\left(\forall y \in X \mid T(b) \neq \sum_{i=1}^m y_i T(a_i)\right) \geq 1 - 2|X|e^{-\mathcal{C}k};$$

for some constant $\mathcal{C} > 0$ (independent of m, k).

1.2 New results

In this paper, we consider the general case where the data set X has no specific structure, and use Gaussian random projections in our arguments to obtain some results about the relationship between ESMP and PESMP.

In the case when X is at most countable (i.e. finite or countable), using a straightforward argument, we prove that these two problems are equivalent almost surely. However, this result is only of theoretical interest due to round-off errors in floating point operations, which make its practical application difficult. We address this issue by introducing a threshold $\delta > 0$ with a corresponding Threshold ESMP (TESMP): if Δ is the distance between $T(p)$ and the closest point of $T(X)$, decide whether $\Delta \geq \delta$.

In the case when X may also be uncountable, we employ the *doubling constant* of X, i.e. the smallest number λ_X such that any closed ball in X can be covered by at most λ_X closed balls of half the radius. Its logarithm $\log_2 \lambda_X$ is called *doubling dimension* of X. Recently, the doubling dimension has become a powerful tool for several classes of problems such as nearest neighbor [10, 8], low-distortion embeddings [3], clustering [12].

We show that we can project X into $\mathbb{R}^k$, where $k = O(\log_2 \lambda_X)$, whilst still ensure the equivalence between ESMP and PESMP with high probability. We also extend this result to the threshold case, and obtain a more useful bound for k.

2 Finite and countable sets

In this section, we assume that X is either finite or countable. Let T be a JL random mapping from a Gaussian distribution, i.e. each entry of T is independently sampled from $\mathcal{N}(0,1)$. It is well known that, for an arbitrary unit vector $a \in \mathbb{S}^{m-1}$, the random variable $\|Ta\|^2$ has a Chi-squared distribution χ_k^2 with k degrees of freedom ([14]). Its corresponding density function is $\frac{2^{-k/2}}{\Gamma(k/2)} x^{k/2-1} e^{k/2}$, where $\Gamma(\cdot)$ is the gamma function. By [4], for any $0 < \delta < 1$, taking $z = \frac{\delta}{k}$ yields a cumulative distribution function

$$F_{\chi_k^2}(\delta) \leq (ze^{1-z})^{k/2} < (ze)^{k/2} = \left(\frac{e\delta}{k}\right)^{k/2}. \tag{3}$$

Thus, we have

$$\mathsf{Prob}(\|Ta\| \leq \delta) = F_{\chi_k^2}(\delta^2) < (3\delta^2)^{k/2} \tag{4}$$

or, more simply, $\mathsf{Prob}(\|Ta\| \le \delta) < \delta^k$ when $k \ge 3$.

Using this estimation, we immediately obtain the following result.

2.1 Proposition
Given $p \in \mathbb{R}^m$ and $X \subseteq \mathbb{R}^m$, at most countable, such that $p \notin X$. Then, for a Gaussian random projection $T : \mathbb{R}^m \to \mathbb{R}^k$ with any $k \ge 1$, we have $T(p) \notin T(X)$ almost surely, i.e. $\mathsf{Prob}\big(T(p) \notin T(X)\big) = 1$.

Proof. First, note that for any $u \neq 0$, $Tu \neq 0$ holds almost certainly. Indeed, without loss of generality we can assume that $\|u\| = 1$. Then for any $0 < \delta < 1$:

$$\mathsf{Prob}\big(T(z) = 0\big) \le \mathsf{Prob}\big(\|Tz\| \le \delta\big) = (3\delta^2)^{k/2} \to 0 \text{ as } \delta \to 0.$$

Since the event $T(p) \notin T(X)$ can be written as the intersection of at most countably many almost sure events $T(p) \neq T(x)$ (for $x \in X$), it follows that $\mathsf{Prob}\big(T(p) \notin T(X)\big) = 1$, as claimed. □

Proposition 2.1 is simple, but it looks interesting because it suggests that we only need to project the data points to a line (i.e. $k = 1$) and study an equivalent membership problem on a line. Furthermore, it turns out that this result remains true for a large class of random projections.

2.2 Proposition
Let ν be a probability distribution on $\mathbb{R}^m$ with bounded Lebesgue density f. Let $Y \subseteq \mathbb{R}^m$ be an at most countable set such that $0 \notin Y$. Then, for a random projection $T : \mathbb{R}^m \to \mathbb{R}^1$ sampled from ν, we have $0 \notin T(Y)$ almost surely, i.e. $\mathsf{Prob}\big(0 \notin T(Y)\big) = 1$.

Proof. For any $0 \neq y \in Y$, consider the set $\mathcal{E}_y = \{T : \mathbb{R}^m \to \mathbb{R}^1 \mid T(y) = 0\}$. If we regard each $T : \mathbb{R}^m \to \mathbb{R}^1$ as a vector $t \in \mathbb{R}^m$, then $\mathcal{E}_y$ is a hyperplane $\{t \in \mathbb{R}^m \mid y \cdot t = 0\}$ and we have

$$\mathsf{Prob}(T(y) = 0) = \nu(\mathcal{E}_y) = \int_{\mathcal{E}_y} f d\mu \le \|f\|_\infty \int_{\mathcal{E}_y} d\mu = 0$$

where μ denotes the Lebesgue measure on $\mathbb{R}^m$. The proof then follows by the countability of Y, similarly to Proposition 2.1. □

Proposition 2.2 is based on the observation that the degree $[\mathbb{R} : \mathbb{Q}]$ of the field extension $\mathbb{R}/\mathbb{Q}$ is $2^{\aleph_0}$, whereas Y is countable; so the probability that any row vector T_i of the random projection matrix T will yield a linear dependence relation $\sum_{j \le m} T_{ij} y_j = 0$ for some $0 \neq y \in Y$ is zero. In practice, however, Y is part of the rational input of a decision problem, and the components of T are rational: hence any subsequence of them is trivially linearly dependent over $\mathbb{Q}$. Moreover, floating point numbers have a bounded binary representation: hence, even if Y is finite, there is a nonzero probability that any subsequence of components of T will be linearly dependent by means of a nonzero multiplier vector in Y.

This idea, however, does not work in practice: we tested it by considering the ESMP given by the IPF defined on the set $\{x \in \mathbb{Z}_+^n \cap [L, U] \mid Ax = b\}$. Numerical experiments indicate that the corresponding PESMP $\{x \in \mathbb{Z}_+^n \cap [L, U] \mid T(A)x = T(b)\}$, with T consisting of a one-row Gaussian projection matrix, is always feasible despite the infeasibility of the original IPF. Since Prop. 2.1

assumes that the components of T are real numbers, we think that the reason behind this failure is the round-off error associated to the floating point representation used in computers. Specifically, when $T(A)x$ is too close to $T(b)$, floating point operations will consider them as a single point. In order to address this issue, we force the projected problems to obey stricter requirements. In particular, instead of only requiring that $T(p) \notin T(X)$, we ensure that

$$\mathsf{dist}(T(p), T(X)) = \min_{x \in X} \|T(p) - T(x)\| > \tau,$$

where dist denotes the Euclidean distance, and $\tau > 0$ is a (small) given constant. With this restriction, we obtain the following result.

2.3 Proposition
Given $\tau, \delta > 0$ and $p \notin X \subseteq \mathbb{R}^m$, where X is a finite set, let

$$d = \min_{x \in X} \|p - x\| > 0.$$

Let $T : \mathbb{R}^m \to \mathbb{R}^k$ be a Gaussian random projection with $k \geq \frac{\log(\frac{|X|}{\delta})}{\log(\frac{d}{\tau})}$. Then:

$$\mathsf{Prob}\big(\min_{x \in X} \|T(p) - T(x)\| > \tau\big) > 1 - \delta.$$

Proof. We assume that $k \geq 3$. For any $x \in X$ we have:

$$\begin{aligned} \mathsf{Prob}\big(\|T(p-x)\| \leq \tau\big) &= \mathsf{Prob}\left(\left\|T\big(\frac{p-x}{\|p-x\|}\big)\right\| \leq \frac{\tau}{\|p-x\|}\right) \\ &\leq \mathsf{Prob}\left(\left\|T\big(\frac{p-x}{\|p-x\|}\big)\right\| \leq \frac{\tau}{d}\right) < \frac{\tau^k}{d^k}, \end{aligned}$$

due to (3). Therefore, by the union bound,

$$\begin{aligned} \mathsf{Prob}\big(\min_{x \in X} \|T(p) - T(x)\| > \tau\big) &= 1 - \mathsf{Prob}\big(\min_{x \in X} \|T(p) - T(x)\| \leq \tau\big) \\ &\geq 1 - \sum_{x \in X} \mathsf{Prob}\big(\|T(p) - T(x)\| \leq \tau\big) > 1 - |X| \left(\frac{\tau}{d}\right)^k. \end{aligned}$$

The RHS is greater than or equal to $1-\delta$ if and only if $\left(\frac{d}{\tau}\right)^k \geq \frac{|X|}{\delta}$, which is equivalent to $k \geq \frac{\log(\frac{|X|}{\delta})}{\log(\frac{d}{\tau})}$, as claimed. $\square$

Note that d is often unknown and can be arbitrarily small. However, if both p, X are integral, then $d \geq 1$ and we can select $k > \frac{\log \frac{|X|}{\delta}}{\log \frac{1}{\tau}}$ in the above proposition.

In many cases, the set X is infinite. We show that when this is the case, we can still overcome this difficulty under some assumptions. In particular, we prove that if $X = \{Ax \mid x \in \mathbb{Z}_+^n\}$ where A is an $m \times n$ matrix with integer coefficients which are all positive in at least one row, then for any bounded vector $b \in \mathbb{Z}^m$ the problem $b \in X$ is equivalent, with high probability, to its projection to a $O(\log n)$-dimensional space. The idea is to separate one positive row and apply random projection to the others.

Formally, let us denote by a^i the i-th row and by a_j the j-th column of A. Assume that all entries in the row a^i is positive and all entries of b are bounded by a constant $B > 0$. Remove the row i from A and b to obtain $\tilde{A} = (a'_1, \ldots, a'_n) \in \mathbb{Z}^{(m-1)\times n}$ and $\tilde{b} \in \mathbb{Z}^{m-1}$. Let $T : \mathbb{R}^{m-1} \to \mathbb{R}^k$ be a JL random mapping and denote by $Z = \{x \in \mathbb{Z}^n_+ \mid a^i \cdot x = b_i\}$. Then we have:

2.4 Proposition
Assume that $b \notin X$, and let $0 < \delta < 1$. Using the terminology and given the assumptions above, if $k \geq \frac{1}{\mathcal{C}} \ln(\frac{2}{\delta}) + \frac{B}{\mathcal{C}} \log(n + B - 1)$ we have

$$\mathsf{Prob}\left(T(b) \neq \sum_{j=1}^{n} x_j T(a'_j) \text{ for all } x \in Z\right) \geq 1 - \delta$$

for some constant $\mathcal{C} > 0$.

Proof. We first show that $|Z| \leq (n + B - 1)^B$. Since all the entries of A are positive integers, we have

$$|Z| \leq |\{x \in \mathbb{Z}^n_+ \mid \sum_{j=1}^{n} x_j = b_i\}| \leq |\{x \in \mathbb{Z}^n_+ \mid \sum_{j=1}^{n} x_j = B\}|.$$

The number of elements in the RHS corresponds to the number of combinations with repetitions of B items sampled from n, which is equal to $\binom{n+B-1}{n-1} = \binom{n+B-1}{B} \leq (n + B - 1)^B$.

Next, by Lemma 1.4, we have:

$$\mathsf{Prob}\left(T(b) \neq \sum_{j=1}^{n} x_j T(a'_j) \text{ for all } x \in Z\right) \geq 1 - 2(n + B - 1)^B e^{-\mathcal{C}k}, \tag{5}$$

which is greater than $1 - \delta$ when taking any k such that $k \geq \frac{1}{\mathcal{C}} \ln(\frac{2}{\delta}) + \frac{B}{\mathcal{C}} \log(n + B - 1)$. The proposition is proved. □

Note that in Prop. 2.4 we can choose the JL random mapping T as a matrix with $\{-1, +1\}$ entries (Rademacher variables). In this case, there is no need to worry about floating point errors.

3 Sets with low doubling dimension

In this section, we denote by $B(x, r)$ the closed ball centered at x with radius $r > 0$, and $B_X(x, r) = B(x, r) \cap X$. We will also assume that X is a doubling space, i.e. a set with bounded doubling dimension. One example of doubling spaces is a Euclidean space. $\mathbb{R}^m$, we can show that the doubling dimension $\log_2(\lambda_X)$ of X can be shown to be a constant factor of m ([16, 6]). However, many sets of low doubling dimensions are contained in high dimensional spaces ([1]). Note that computing the doubling dimension of a metric space is generally **NP**-hard ([5]). We shall make use of the following simple lemma.

3.1 Lemma
For any $p \in X$ and $\varepsilon, r > 0$, there is a set $S \subseteq X$ of size at most $\lambda_X^{\lceil \log_2(\frac{r}{\varepsilon}) \rceil}$ such that

$$B_X(p, r) \subseteq \bigcup_{s \in S_j} B(s, \varepsilon).$$

Proof. By definition of the doubling dimension, $B_X(p,r)$ is covered by at most λ_X closed balls of radius $\frac{r}{2}$. Each of these balls in turn is covered by λ_X balls of radius $\frac{r}{4}$, and so on: iteratively, for each $k \geq 1$, $B_X(p,r)$ is covered by λ_X^k balls of radius $\frac{r}{2^k}$. If we select $k = \lceil \log_2(\frac{r}{\varepsilon}) \rceil$ then $k \geq \log_2(\frac{r}{\varepsilon})$, i.e. $\frac{r}{2^k} \leq \varepsilon$. This means $B_X(p,r)$ is covered by $\lambda_X^{\lceil \log_2(\frac{r}{\varepsilon}) \rceil}$ balls of radius ε. □

We will also use the following lemma, which is proved in [8] using a concentration estimation for sum of squared gaussian variables (Chi-squared distribution).

3.2 Lemma
Let $X \subseteq B(0,1)$ be a subset of the m-dimensional Euclidean unit ball. Then there exist universal constants $c, C > 0$ such that for $k \geq C \log \lambda_X + 1$ and $\delta > 1$, the following holds:

$$\mathsf{Prob}(\exists x \in X \text{ s.t. } \|Tx\| > \delta) < e^{-ck\delta^2}.$$

In the proof of the next result (one of the main results in this section), we use the same idea as that in [8] for the nearest neighbor problem.

3.3 Theorem
Given $0 < \delta < 1$ and $p \notin X \subseteq \mathbb{R}^m$. Let $T : \mathbb{R}^m \to \mathbb{R}^k$ be a Gaussian random projection. Then

$$\mathsf{Prob}(T(p) \notin T(X)) = 1$$

if $k \geq \mathcal{C} \log_2(\lambda_X)$, for some universal constant $\mathcal{C}$.

Proof. Let $\varepsilon > 0$ and $0 = r_0 < r_1 < r_2 < \dots$ be positive scalars (their values will be defined later). For each $j = 1, 2, 3, \dots$ we define a set

$$X_j = X \cap B(p, r_j) \setminus B(p, r_{j-1}).$$

Since $X_j \subseteq B_X(p, r_j)$, by Lemma 3.1 we can find a point set $S_j \subseteq X$ of size $|S_j| \leq \lambda_X^{\lceil \log_2(\frac{r_j}{\varepsilon}) \rceil}$ such that

$$X_j \subseteq \bigcup_{s \in S_j} B(s, \varepsilon).$$

Hence, for any $x \in X_j$, there is $s \in S_j$ such that $\|x - s\| < \varepsilon$. Moreover, by the triangle inequality, any such s satisfies $r_{j-1} - \varepsilon < \|s - p\| < r_j + \varepsilon$, so without loss of generality we can assume that

$$S_j \subseteq B(p, r_j + \varepsilon) \setminus B(p, r_{j-1} - \varepsilon).$$

We denote by $\mathcal{E}_j$ the event that:

$$\exists s \in S_j,\ \exists x \in X_j \cap B(s, \varepsilon) \text{ s.t. } \|Ts - Tx\| > \varepsilon\sqrt{j}.$$

By the union bound, we have

$$\begin{aligned}
\mathsf{Prob}(\mathcal{E}_j) &\leq \sum_{s \in S_j} \mathsf{Prob}\big(\exists x \in X_j \cap B(s, \varepsilon) \text{ s.t. } \|Ts - Tx\| > \varepsilon\sqrt{j}\big) \\
&\leq \sum_{s \in S_j} e^{-c_1 kj} \qquad \text{(for some universal constant } c_1 \text{ by Lemma 3.2)} \\
&\leq \lambda_X^{\lceil \log_2(\frac{r_j + \varepsilon}{\varepsilon}) \rceil} e^{-c_1 kj}.
\end{aligned}$$

Again by the union bound, we have:

$$\begin{aligned}\mathsf{Prob}\big(\exists x \in X \text{ s.t } T(x) = T(p)\big) &= \mathsf{Prob}\big(\exists x \in \bigcup_{j=1}^{\infty} X_j \text{ s.t } T(x) = T(p)\big) \\ &\leq \sum_{j=1}^{\infty} \mathsf{Prob}\big(\exists x \in X_j \text{ s.t } T(x) = T(p)\big).\end{aligned}$$

Now we will estimate the individual probabilities:

$$\begin{aligned}&\mathsf{Prob}\big(\exists x \in X_j \text{ s.t } T(x) = T(p)\big) \\ \leq\ &\mathsf{Prob}\big((\exists x \in X_j \text{ s.t } T(x) = T(p)) \wedge \mathcal{E}_j^c\big) + \mathsf{Prob}(\mathcal{E}_j) \\ \leq\ &\mathsf{Prob}\big(\exists x \in X_j, s \in S_j \cap B(x,\varepsilon) \text{ s.t } T(x) = T(p) \wedge \|T(s) - T(x)\| \leq \varepsilon\sqrt{j}\big) + \mathsf{Prob}(\mathcal{E}_j) \\ \leq\ &\mathsf{Prob}\big(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| < \varepsilon\sqrt{j}\big) + \lambda_X^{\lceil \log_2(\frac{r_j+\varepsilon}{\varepsilon})\rceil} e^{-c_1 kj}.\end{aligned}$$

Next, we choose $\varepsilon = \frac{d}{N}$ for some large N; and for each $j \geq 1$, we choose $r_j = (2+j)\varepsilon$. For $j < N-2$, by definition it follows that $X_j = \varnothing$. Therefore

$$\mathsf{Prob}\big(\exists x \in X_j \text{ s.t } T(s) = T(p)\big) = 0.$$

On the other hand, for $j \geq N-2$,

$$\begin{aligned}&\mathsf{Prob}\big(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| \leq \varepsilon\sqrt{j}\big) \\ \leq\ &\lambda_X^{\lceil \log_2(\frac{r_j+\varepsilon}{\varepsilon})\rceil} \mathsf{Prob}\big(\|T(z)\| \leq \frac{\varepsilon\sqrt{j}}{r_{j-1} - \varepsilon}\big) \quad \text{for an arbitrary } z \in \mathbb{S}^{n-1} \\ =\ &\lambda_X^{\lceil \log_2(3+j)\rceil} \mathsf{Prob}\big(\|T(z)\| \leq \frac{1}{\sqrt{j}}\big) \quad \text{for an arbitrary } z \in \mathbb{S}^{n-1} \\ <\ &\lambda_X^{\lceil \log_2(3+j)\rceil} j^{-k/2} \qquad \text{by the estimation (4).}\end{aligned}$$

Note that $\lambda_X^{\lceil \log_2(3+j)\rceil} \leq \lambda_X^{\log_2(6+2j)} = (6+2j)^{\log_2 \lambda_X} < j^{(2\log_2 \lambda_X)}$ for large enough N. Therefore, we have

$$\begin{aligned}\mathsf{Prob}\big(\exists x \in X_j \text{ s.t } T(x) = T(p)\big) &\leq \lambda_X^{\lceil \log_2(3+j)\rceil}\big(j^{-k/2} + e^{-c_1 kj}\big) \\ &\leq j^{-c_2 k} + e^{-c_3 kj}\end{aligned}$$

for some universal constants c_2, c_3, provided that $k \geq \mathcal{C}_1 \log \lambda_X$ for some large enough constant $\mathcal{C}_1$. Finally, by the union bound,

$$\begin{aligned}\mathsf{Prob}\big(T(p) \notin T(X)\big) &= 1 - \mathsf{Prob}\big(T(p) \in T(X)\big) \\ &\geq 1 - \sum_{i=N-2}^{\infty} \big(i^{-c_2 k} + e^{-c_3 kj}\big)\end{aligned}$$

which tends to 1 when N tends to infinity. $\square$

Our final result in the section is an extension of Thm. 3.3 to the threshold case.

3.4 Theorem
Let $p \notin X \subseteq \mathbb{R}^m$, $T : \mathbb{R}^m \to \mathbb{R}^k$ be a Gaussian random projection, and $d = \min_{x \in X} \|p - x\|$. Then for all $0 < \delta < 1$ and all $0 < \tau < \kappa d$ for some constant $\kappa < 1$, we have

$$\mathsf{Prob}(\mathsf{dist}(T(p), T(X)) > \tau) > 1 - \delta$$

if k is $O(\frac{\log(\frac{\lambda_X}{\delta})}{\log(\frac{d}{\tau})})$.

Proof. For $j = 1, 2, \ldots$ we construct the sets X_j, S_j similarly as those in the proof of Thm. 3.3 (where the values of r_j and ε will be defined later). Then we have

$$\begin{aligned}\mathsf{Prob}\big(\exists x \in X \text{ s.t } \|T(x) - T(p)\| < \tau\big) &= \mathsf{Prob}\big(\exists x \in \bigcup_{j=1}^{\infty} X_j \text{ s.t } \|T(x) - T(p)\| < \tau\big) \\ &\leq \sum_{j=1}^{\infty} \mathsf{Prob}\big(\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau\big).\end{aligned}$$

For all $j \geq 1$, we have

$$\begin{aligned}&\mathsf{Prob}\big(\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau\big) \\ \leq\ &\mathsf{Prob}\big((\exists x \in X_j \text{ s.t } \|T(x) - T(p)\| < \tau) \wedge \mathcal{E}_j^c\big) + \mathsf{Prob}(\mathcal{E}_j) \\ \leq\ &\mathsf{Prob}\big(\exists x \in X_j, s \in S_j \cap B(x, \varepsilon) \text{ s.t } \|T(x) - T(p)\| < \tau \wedge \|T(s) - T(x)\| \leq \varepsilon\sqrt{j}\big) + \mathsf{Prob}(\mathcal{E}_j) \\ \leq\ &\mathsf{Prob}\big(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| < \tau + \varepsilon\sqrt{j}\big) + \lambda_X^{\lceil \log_2(\frac{r_j + \varepsilon}{\varepsilon})\rceil}\ e^{-c_1 k j}.\end{aligned}$$

Now we choose $\varepsilon = \frac{\tau}{N}$ for some $N > 0$ such that $1 + \frac{1}{N} < \frac{1}{\kappa}$ and for each $j \geq 1$, we choose $r_j = \tau\sqrt{j+1} + (2 + j)\varepsilon$. For $j = 1$, by the union bound we have

$$\begin{aligned}&\mathsf{Prob}\big(\exists s \in S_1 \text{ s.t } \|T(s) - T(p)\| \leq \tau + \varepsilon\sqrt{1}\big) \\ \leq\ &\lambda_X^{\lceil \log_2(\frac{r_1 + \varepsilon}{\varepsilon})\rceil} \mathsf{Prob}\big(\|T(z)\| \leq \frac{\tau + \varepsilon}{d}\big) \quad \text{for an arbitrary } z \in \mathbb{S}^{m-1} \\ =\ &\lambda_X^{\lceil \log_2(4 + N\sqrt{2})\rceil} \mathsf{Prob}\left(\|T(z)\| \leq (1 + \frac{1}{N})\frac{\tau}{d}\right) \quad \text{for an arbitrary } z \in \mathbb{S}^{m-1} \\ <\ &\lambda_X^{\lceil \log_2(4 + N\sqrt{2})\rceil} \left((1 + \frac{1}{N})\frac{\tau}{d}\right)^{k/2} \quad \text{by estimation (4)} \\ <\ &\left((1 + \frac{1}{N})\frac{\tau}{d}\right)^{c_2 k}\end{aligned} \tag{6}$$

for some universal constant $c_2 > 0$, as long as $k > \mathcal{C}\log(\lambda_X)$ for some $\mathcal{C}$ large enough.

For $j \geq 2$, we have

$$\begin{aligned}&\mathsf{Prob}\big(\exists s \in S_j \text{ s.t } \|T(s) - T(p)\| \leq \tau + \varepsilon\sqrt{j}\big) \\ \leq\ &\lambda_X^{\lceil \log_2(\frac{r_j + \varepsilon}{\varepsilon})\rceil} \mathsf{Prob}\big(\|T(z)\| \leq \frac{\tau + \varepsilon\sqrt{j}}{r_{j-1} - \varepsilon}\big) \quad \text{for an arbitrary } z \in \mathbb{S}^{m-1} \\ =\ &\lambda_X^{\lceil \log_2(3 + j + N\sqrt{j+1})\rceil} \mathsf{Prob}\big(\|T(z)\| \leq \frac{1}{\sqrt{j}}\big) \quad \text{for an arbitrary } z \in \mathbb{S}^{m-1} \\ <\ &\lambda_X^{\lceil \log_2(3 + j + N\sqrt{j+1})\rceil}\ j^{-k/2} \quad \text{by estimation (4)} \\ <\ &j^{-c_3 k}\end{aligned} \tag{7}$$

for some universal constant $c_3 > 0$, as long as $k > \mathcal{C}\log(\lambda_X)$ for some $\mathcal{C}$ large enough.

Similarly, for all $1 \leq j$, we have

$$\lambda_X^{\lceil \log_2(\frac{r_j+\varepsilon}{\varepsilon})\rceil} e^{-c_1 kj} \leq e^{-c_4 kj}, \tag{8}$$

for some universal constant $c_4 > 0$, as long as $k > \mathcal{C}\log(\lambda_X)$ for some $\mathcal{C}$ large enough.

From estimations (6), (7), (8) and by the union bound we have:

$$\begin{aligned}
\mathsf{Prob}(\mathsf{dist}(T(p), T(X)) \geq \tau) &\geq 1 - \sum_{j=1}^{\infty} \mathsf{Prob}(\mathsf{dist}(T(p), T(X_j)) < \tau) \\
&\geq 1 - \left((1 + \frac{1}{N})\frac{\tau}{d}\right)^{c_2 k} - \sum_{j=2}^{\infty} j^{-c_3 k} - \sum_{j=1}^{\infty} e^{-c_4 kj} \\
&\geq 1 - \delta \qquad \text{for } k = O(\tfrac{\log(\frac{\lambda_X}{\delta})}{\log(\frac{d}{\tau})}) \text{ large enough.}
\end{aligned}$$

□

References

[1] I. Abraham, Y. Bartal, and O. Neiman. Embedding metric spaces in their intrinsic dimension. In *19th Annual ACM-SIAM Symposium on Discrete Algorithms*, page 363372, 2008.

[2] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66:671–687, 2003.

[3] P. Agarwal, S. Har-Peled, and H. Yu. Embeddings on surfaces, curves, and moving points in Euclidean space. In *Proceedings of the 23rd Symposium on Computational Geometry*, pages 381–389. ACM, 2007.

[4] S. Dasgupta and A. Gupta. An elementary proof of a theorem by johnson and lindenstrauss. *Random Structures and Algorithms*, 22:60–65, 2002.

[5] L. Gottlieb and R. Krauthgamer. Proximity algorithms for nearly doubling spaces. *SIAM Journal on Discrete Mathematics*, 27(4):1759–1769, 2013.

[6] A. Gupta, R. Krauthgamer, and J.R. Lee. Bounded geometries, fractals, and low-distortion embeddings. In *IEEE Symposium on Foundations of Computer Science*, page 534543. IEEE, 2003.

[7] P. Indyk. Algorithmic applications of low-distortion geometric embeddings. In *42nd Annual Symposium on Foundations of Computer Science*, pages 10–33. IEEE, 2001.

[8] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Transactions on Algorithms*, 3(3):Art. 31, 2007.

[9] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In G. Hedlund, editor, *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, Providence, 1984. AMS.

[10] R. Krauthgamer and J.R. Lee. Navigating nets: Simple algorithms for proximity search. In *Proceedings of the 15th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 791–801, 2004.

[11] Vu Khac Ky, P.-L. Poirion, and L. Liberti. Using the Johnson-Lindenstrauss lemma in linear and integer programming. Technical Report 07/4985, Optimization Online, 2015.

[12] A. Magen. Dimensionality reductions in ℓ_2 that preserve volumes and distance to affine spaces. *Discrete and Computational Geometry*, 30(1):139–153, 2007.

[13] J. Matousek. *Lectures on Discrete Geometry*. Springer, New York, 2002.

[14] A. Mood, F. Graybill, and D. Boes. *Introduction to the Theory of Statistics*. McGraw-Hill, 1974.

[15] K. Murty and S. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.

[16] J.-L. Verger-Gaugry. Covering a ball with smaller equal balls in $\mathbb{R}^n$. *Discrete Computational Geometry*, 33:143–155, 2005.